Grid Applicability to the Earth Sciences

February 18, 2004

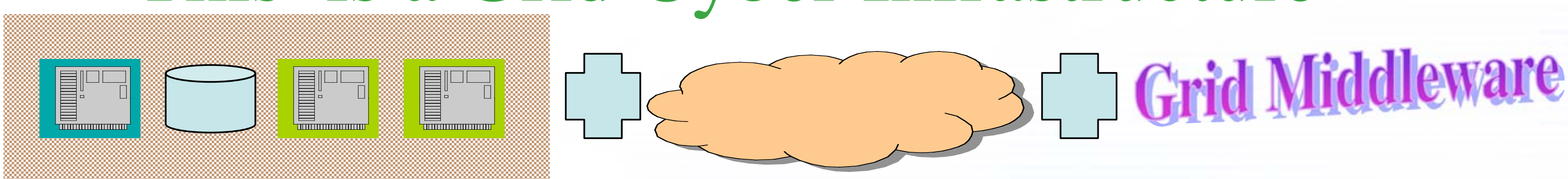
Outline

- What is a grid?
- Why grid technology is important to NASA
- Current status of grids at NASA
- Current status of NASA Earth science-related grid applications
- Roadmap to an Earth Science grid
- Conclusions

What Do Grids Do?

- Grids provide the intelligent cyber infrastructure
  - To dynamically integrate independently managed:
    - Computational resources
    - Data archives
    - Scientific Sensors/Instruments
  - To build large scale collaborative problem solving environments that are:
    - Cost effective
    - Secure
- Grid software is “middleware”

This is a Grid Cyber Infrastructure



Resources

Networks

Why Grid Technology Is Important to NASA?

- Supports single sign on to allow users to securely used many distributed resources (processors and archives) without having to individually authenticate to each -- the grid handles this.
- Allows users to easily use data and computational resources, irrespective of their location within the NASA community.
- Supports grid-enabled, self-describing services that can be easily tied into complex processing workflows.
- Is part of an international collaboration/standardization movement
  - Centered around the Global Grid Forum
  - Involving significant investments by the US (DOE and NASA), European Union, UK, Korea, Japan, China
  - Involving a growing list of vendors

Current Status of Grids at NASA

- NASA has developed the Information Power Grid that currently encompasses computational resources at
  - Ames Research Center, Glenn Research Center, Langley Research Center, JPL, with planning underway for Goddard Space Flight Center, Marshall Space flight Center, Johnson Space Center
- NASA is developing grid technology to increase the intelligence of the grid
- Since NASA may ultimately have many grids, ARC is:
  - providing technical support to other NASA organizations that are interested in deploying grid technology
  - deploying ARC-developed grid technology across the various NASA grids that may be developed

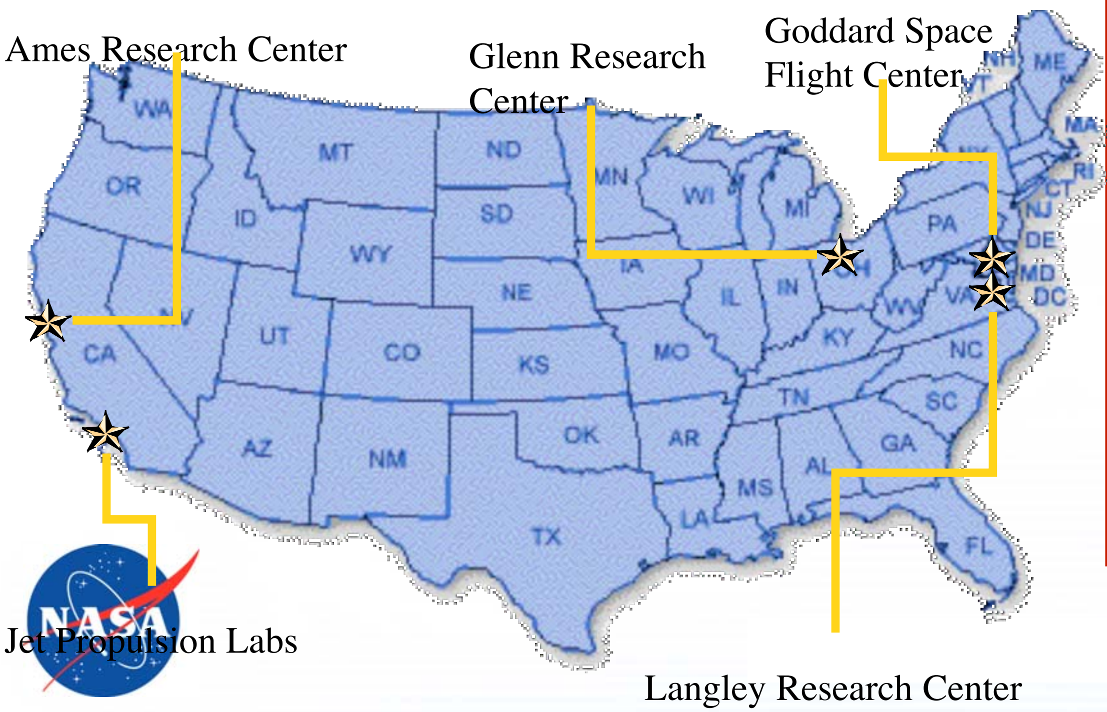
NASA ADVANCED SUPERCOMPUTING

ARC Information Power Grid (IPG)

Vision:

To make the practice of large-scale science and engineering, as well as other widely distributed, data intensive NASA activities, much more effective than it is today.

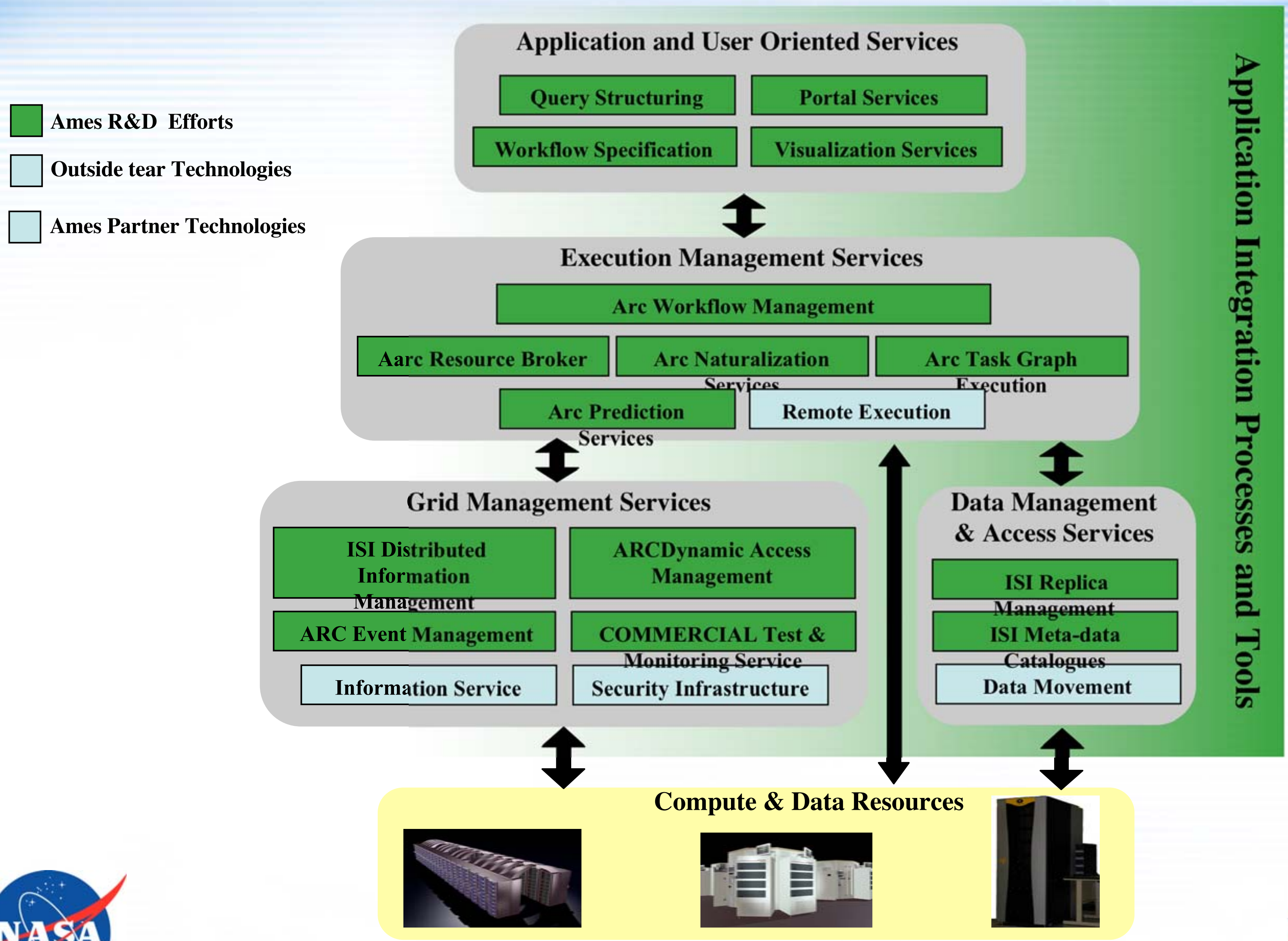
Grid technology is the foundation to making this vision a success



- Grid Nodes
  1. 1024 CPU SGI O3K IRIX ARC
  2. 512 CPU SGI O3K IRIX ARC
  3. 512 CPU SGI Altrix LINUX ARC (in progress)
  4. 128 CPU (node) Linux Cluster GRC
  5. 128 CPU SGI O2K IRIX ARC
  6. 128 CPU SGI O2K IRIX ARC
  7. 64 CPU SIG O2K IRIX ARC
  8. 32 CPU SGI O2K IRIX ARC
  9. 24 CPU SGI O3K IRIX ARC
  10. 24 CPU SGI O2K IRIX GRC
  11. 16 CPU SGI O2K IRIX LaRC
  12. 16 CPU SGI O2K IRIX ARC
  13. 12 CPU (node) Linux Cluster (in progress)
  14. 8 CPU SGI O3K IRIX LaRC
  15. 8 CPU SUN Ultra Sparc3 ARC - Storage
  16. 8 CPU SGI O2K IRIX ARC - Storage
  17. 8 CPU SGI O3K IRIX JPL
  18. 8 CPU SGI O3K IRIX GSFC (planned)
  19. 4 CPU LINUX ARC (planned)

NAS
NASA Advanced Supercomputing

Grid Middleware Services



Status: Work to Facilitate Efficient Use of Distributed Computational Resources

- Ongoing NASA work to increase the intelligence of grid processing
  - Resource discovery through brokers that select the “best” set of resources based on user requirements
  - Execution managers that can autonomously manage the user’s job as it moves through stages of grid processing
  - Dynamic access services that permit users to instantly, but with proper accountability, access needed computational resources across administrative boundaries, without having accounts on these machines
  - Naturalization service that automatically tailors the processing environment on grid resources

Status: Work to Facilitate Utilization of Distributed Heterogeneous Data

- Ongoing NASA work to increase the intelligence of grid data handling
  - Data discovery through distributed, grid-accessible metadata catalogs and efficient access through replicated grid data caches.
  - Efficient grid-enabled access to data stored on secondary and tertiary storage devices.
  - On-demand data product generation and access though
    - Virtual data object (store recipe for high level data product not data itself)
    - Treat Virtual data object as first class object (same as stored data)

Status: Work to Support Grid Cyber Infrastructure Management

- Ongoing work to increase the intelligence of grid management and assessment
  – Develop tools and processes that enable rapid integration of grid technologies for NASA applications.
  – Develop software that enables rapid installation, deployment, maintenance and monitoring of distributed services across heterogeneous and federated sets of resources.
  – Develop information services that accurately and quickly provide information about grid resources

Status: Grid Service Architectures

- The grid community and the web service community are collaborating to provide a service architecture
  - Remote access to data by software, not human access is the goal of web page technology
    - Web has to do with human access to data using web browsers.
    - Web services has to do with software access to data.
  - Wraps data in the XML markup language
  - Is intended for loosely coupled distributed systems in contrast to CORBA and EJB (Enterprise Java Beans) which are intended for tightly coupled distributed systems.
  - Service described through Web Service Description Language (WSDL)

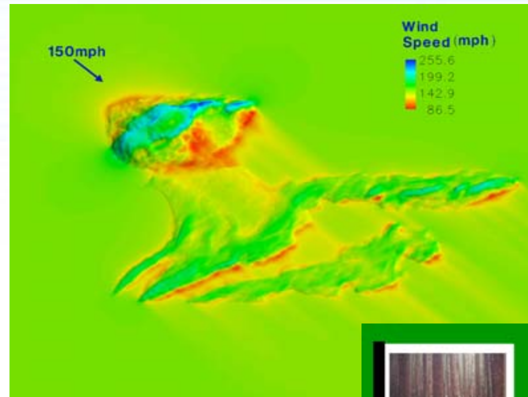
NASA Grid Support for Earth Science

- NASA grid work has supported Earth Science applications
  - ARC Forest Vegetation Dynamics (Marc Kramer)
  - JPL QuakeSim (Andrea Donnellan)
  - George Mason University/ARC Integration of Open GIS Consortium and Grid Technology for Earth Science Modeling and Applications under NASA AIST program (Liping Di)
  - LaRC Atmospheric Sciences Data Center Distributed Generation of Earth Science Data Products Using Grids (Bruce Barkstrom)
  - CEOS Grid Testbed
- NASA grid work with other Non-Earth Science applications may be applicable to Earth Science problems.
  - ARC parameter study applications
  - MSFC/JSC International Space Station Experiments

NASA ADVANCED SUPERCOMPUTING

Forest Vegetation Dynamics

Goals and Objectives: To process and integrate real-time EOS data from different instruments with output from a computational fluid dynamics (CFD) model in order to study the effects of extreme weather on carbon amounts and fluxes at the top of a forest canopy



Current Status:

- Demonstrated grid-enabled, scaleable workflow, coupling data processing/fusion, CFD modeling, and linear airflow dynamics
- Processed LiDAR, SRTM, and Landsat Thematic-Mapper (TM) data on three different grid compute nodes, combined the LiDAR and TM data on a fourth, and stored the processed data on a single IPG mass-storage device

Future Plans:

- Process larger data sets for additional CFD simulations

Contribution To NASA Grid Technology:

Demonstrated the value of grid technology to a next-generation Earth Science application

Participants: ARC

QuakeSim

Goals and Objectives: To develop a grid-enabled geophysics problem-solving environment (PSE) for studying active tectonic and earthquake processes

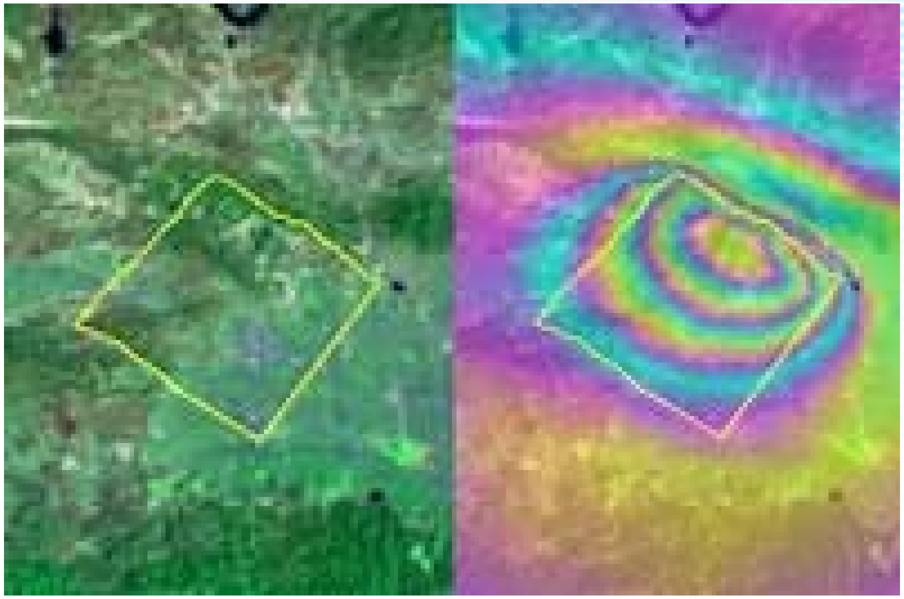
- Simulation, data manipulation, pattern recognition, and analysis tools must interact with data, both real and generated, and run in a heterogeneous, distributed computing environment

General Architecture:

- Makes use of Grid Services
- Interfaces software through an object grid framework whose middleware is built on Java servers linked by SOAP (Simple Object Access Protocol) and RMI (Remote Method Invocation)
- Defines distributed object API's (Application Programming Interface) in XML (Extensible Markup Language) to support interoperability.

Current Status:

- Object grid framework currently under development and scheduled for completion in late FY04
- GeoFEST application code successfully ported to the NAS SGI Altix (A single image 512 processor Linux machine)
- Virtual California currently undergoing testing and benchmarking on NAS IPG machines



Contribution To NASA Grid Technology:

QuakeSim will be the first science application to exploit Grid Service technology on the IPG.

Participants: JPL, ARC, Goddard, Indiana University, University of California

NAS
NASA Advanced Supercomputing

Integration of OGC and Grid Technologies for Earth Science Modeling and Applications

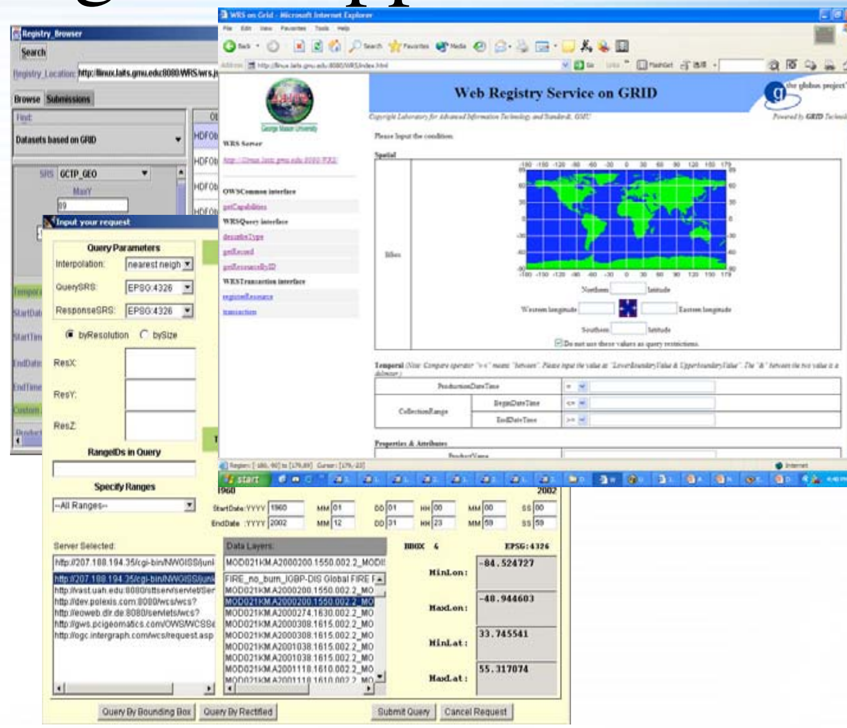
Goals and Objectives: Making NASA EOSDIS data easily accessible to Earth science modeling and applications communities by combining the advantages of both Open GIS (Geographical Information System) Consortium (OGC) and Grid technology thereby achieving the integration in NASA EOSDIS Data Pools and ESG (Earth System Grid).

Current Status:

- Established the testbed environments:
  - Installed Grid software at George Mason University (GMU) Laboratory for Advanced Information Technology & Standards (LAITS) and NASA Ames.
- Provided multiple versions of GMU-developed OGC compliant software for various Operating Systems (Linux, IRIX)
- Designed and implemented the Spatial Grid Service (SGS) as an agent between OGC compliant software and Grid software.

Future Plans:

- Design and implement a Grid Services-based Spatial Data Services with Globus toolkit 3.0.
- Design a metadata catalogue for data replica catalogue and produce a catalogue architecture scalable to Data Pools



Contribution To NASA Grid Technology:

Driving the use of data grid services within the EOSDIS community to provide a geospatially aware data searching tool that spans multiple data repositories.

Participants: George Mason University, ARC

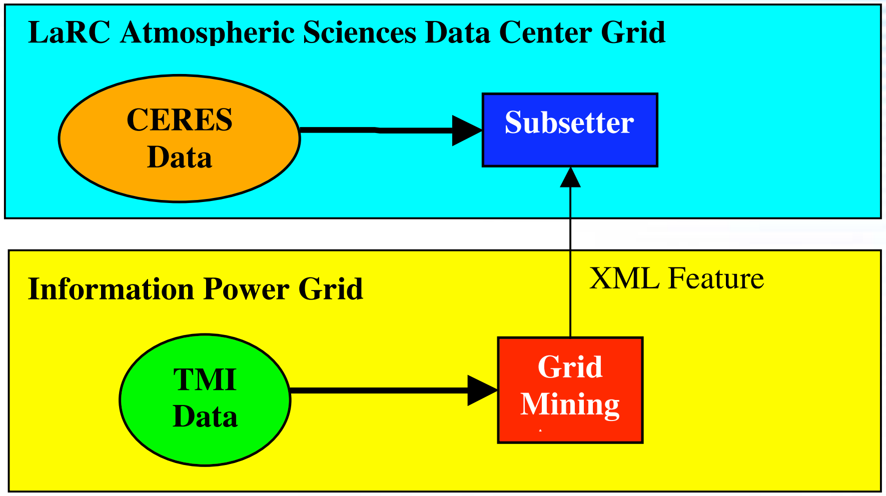
Distributed Generation of Earth Science Data Products Using Grids

Goals and Objectives: Demonstrate the use of grid technology to support both intra-center and inter-center generation of Earth Science data products.

Current Status:
Successfully satisfied June 2003 program milestone:

- Used grid technology to develop subsetter for CERES data at LaRC Atmospheric Sciences Data Center.
- Used Grid Miner to mine GSFC TRMM/TMI data (that was cached at on ARC mass storage system) for mesoscale convective systems (MCS)
- Used grid-enabled Storage Resource broker to extract data directly from mass storage system for use by miner.
- Used grid to transfer MCS feature index (circumscribing polygon) to LaRC to subset CERES data that corresponded to MCS.

Future Plans: Searching for Earth Science Information Partners (ESIPS) that might be interested in working with grid technology.



Contribution To NASA Grid Technology:

Using grid technology to generate a data product that involved data from two NASA centers.

Participants: Atmospheric Sciences Data Center at LaRC, ARC

Committee on Earth Observation Satellites (CEOS) Grid Testbed

Goals and Objectives: The CEOS GRID Task Team will develop technical requirements and identify GRID technologies and services to be implemented in testbed locations.

Current Status:

- ARC supported CEOS Grid Workshop in April 2002 and as a result of the workshop the Working Group on Information Systems Services (WGISS) decided to propose CEOS Grid Testbed
- Go ahead given by CEOS in Sept 2002
- CEOS Grid Testbed participants began work in 2002 and have grown in number since then.
- ARC provided grid consulting as well as host and user grid certifications
- A CEOS Grid Toolkit has been compiled and a Grid monitoring system is in place to monitor testbed systems.

Future Plans: Continue investigating Grid technology. Current interest is in workflow management and data catalogues.

Participants:

- EOSDIS & George Mason University (GMU)
- European Space Agency (ESA)
- DutchSpace
- NOAA Operational Model Archive & Distribution System (NOMADS)
- University of Alabama – Huntsville (UHA)
- United States Geological Survey – EROS Data Center (EDC)
- NASA Advanced Data Grid (ADG)
- China Spatial Information Grid (SIG)
- ARC (supplying host and user certificates and grid consulting)

Contribution To NASA Grid Technology:

Exploring the advantages of Grid technology for Earth Science.

Participants: CEOS Member sites

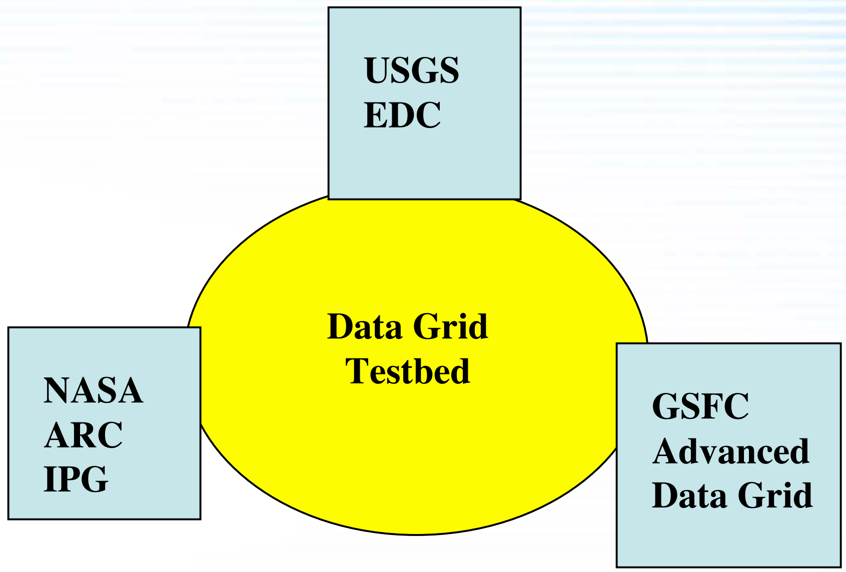
CEOS Participant: USGS EROS Data Center, NASA Advanced Data Grids, ARC IPG

Goals and Objectives: Establish a data grid between USGS EDC, NASA Goddard ADG and IPG.

Current Status:

- These groups are currently in the process of defining the roles at each site and identifying the services each will provide to begin the first phase.
- In phase one, a simulated data fusion algorithm will be initiated at a scientist workstation, processing will take place at another site.
- Data will reside in two locations and be moved using grid technology.

Future Plans: In the second phase, the work will be with the real data fusion algorithm with plans to use the IPG to demonstrate processing these data sets and to experiment with grid workflow techniques.



Contribution To NASA Grid Technology:

Develop methods for close collaboration among geographically distributed locations using Grid Technology.

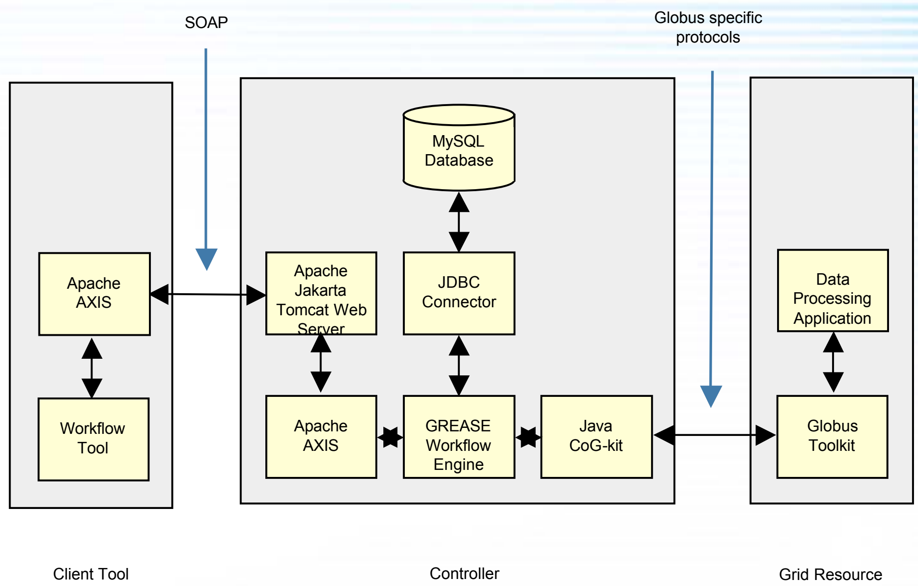
Participants: GSFC ADG, USGS EDC, ARC IPG

CEOS Participant: Grid Aware End-to-end Analysis and Simulation Environment

Goals and Objectives: Create a software environment where simulators of EO instruments and data processing software can work together using computational grid technology.

Current Status:
Workflow management application successfully completed:
- Multiple distributed compute and data storage resources may be used
- Applications executed at distributed locations
- Geographically distributed simulations
- Start jobs and monitor progress behind firewalls
- Provides easy to use interfaces
- Controller implemented as Web Service
  - Uses SOAP
  - Description via WSDL
- Graphical user interface
  - Implemented in Java

Future Plans: Provide support for Globus Toolkit 3, dynamic resource management, dynamic application staging, and advanced control structures. Name change to GridAssist.



Contribution To NASA Grid Technology:

Provides a tool to allow users to make use of the grid without having specific Grid expertise.

Participants: DutchSpace, CEOS, European Space Agency (ESA)

NASA ADVANCED SUPERCOMPUTING

CEOS Participant: Grid Engine

Goals and Objectives: Develop an Open Grid service for Earth Observation (EO) infrastructure where specific data handling and application services are seamlessly plugged-in.

Current Status: Grid Engine framework is in development. The initial system perform these tasks:

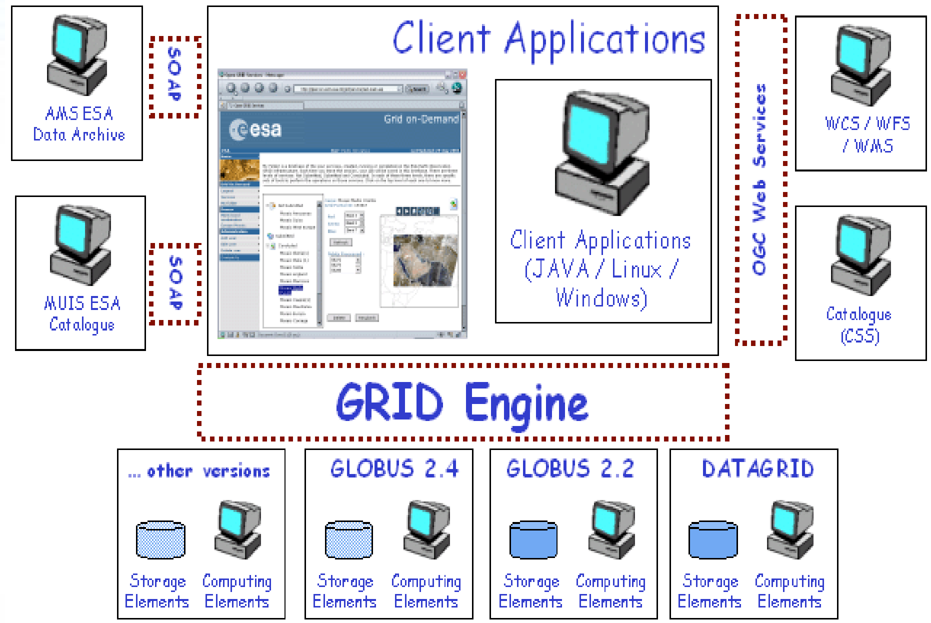
- Connect and create secure Grid Sessions
- File upload and download using GridFTP
- Simple job submission in multiple Grid environments
- Session and basic job submission status

• EO Grid interfaces based on web services

- Allows access and management of multiple grids

• Demo with ENVISAT MERIS data

Future Plans: Build interface to MODIS. Integrate the EO in the Datagrid Replica Metadata Service. Begin work with DutchSpace and China-SIG



Contribution To NASA Grid Technology:

Expected to provide the flexibility for building an application virtual community with quick accessibility to data, compute resources and results.

Participants: ESA-ESRIN

CEOS Participant: China Spatial Information Grid (SIG)

Goals and Objectives: Build a spatial information sharing and spatial resource cooperation platform.

Current Status:

A test grid, T-SIG, was built connecting China Remote Sensing Ground Station (RSGS-CHINA), Tsinghua University, and Chinese Academy of Forestry (CAF-CHINA)

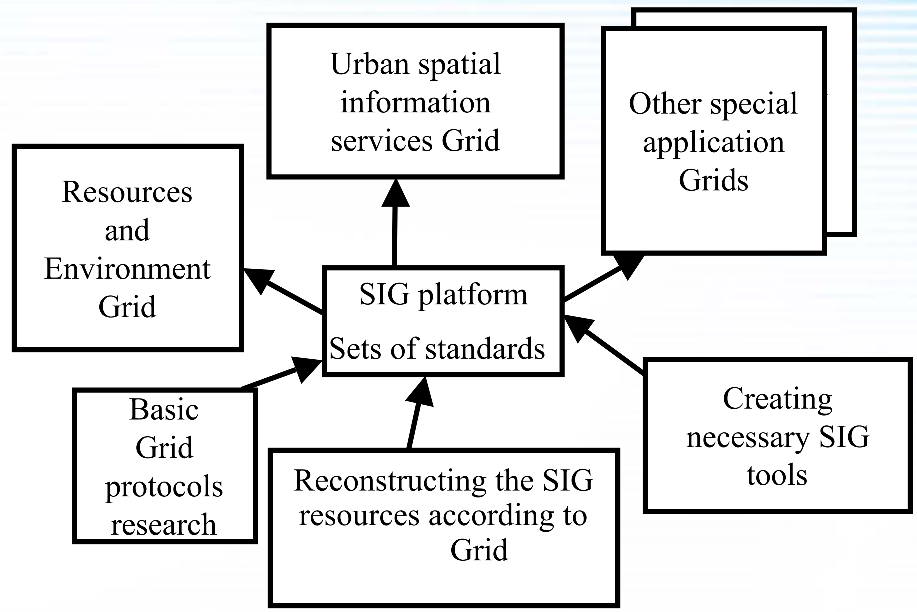
Registered resources include two remote sensing image databases, a CAF forestry GIS database, PCI desktop software, and GeoMedia GIS software.

Work completed includes:

- EO Data Viewer based for remote sensing data
- Visualization Interface
- A description language GAPL and its runtime environment GARE for applications
- Dedicated spatial resource register using P2P technology

Future Plans: Increase compounds of SIG, and obtain high performance for applications.

SIG Structure

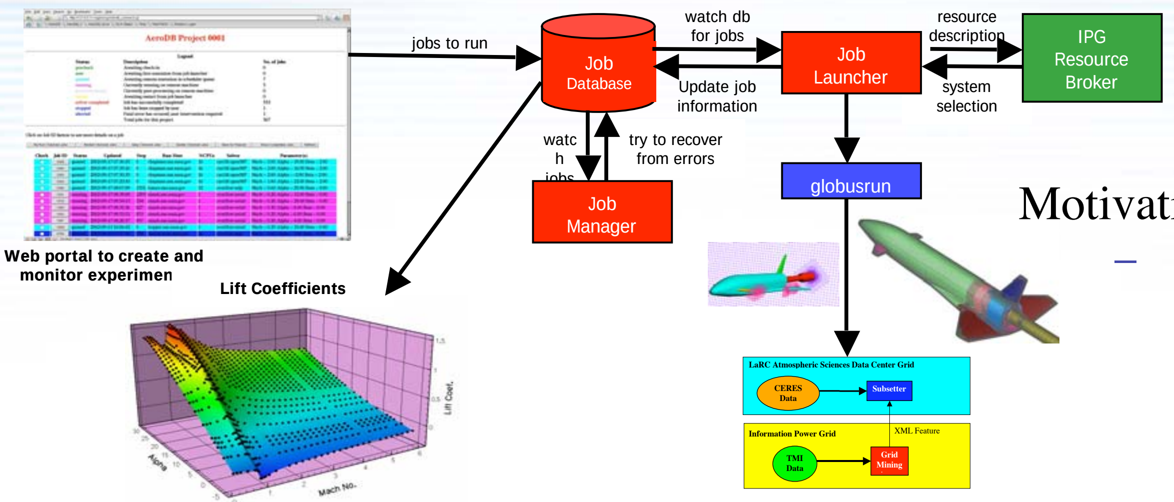


Contribution To NASA Grid Technology:

Development of Earth Science tools.

Participants: CEOS, China SIG

Parameter Studies on The Grid
AeroDB Overview



Motivation

- To build a Stability and Control database, one might require:
  - 10 angles of attack
  - 5 side slip angles
  - 25 Mach numbers
  - 10 aileron deflections
  - 10 rudder deflections
  - 10 elevator deflections
- Total of 1.25 million conditions

Goal:

- Automate the process of running thousands of CFD cases with little user intervention
- Use the NASA IPG

Run a large CFD parameter study

Result

Location	Host	Hardware Type	Number of jobs	Approx CPU Hours
ARC	hostA	SGI O3K	3489	25485.0
	hostB	SGI O3K	1074	15678.3
	hostC	SGI O2K	477	8017.0
	hostD	SGI O2K	411	4702.3
	hostE	SGI O2K	61	262.3
	hostF	Sun Ultra	136	234.6
GRC	hostF	SGI O2K	126	1014.1
	hostG	Linux PC Cluster	70	976.1
NCSA	hostH	SGI O2K	99	483.3
ISI	hostI	SGI O2K	21	212.3
Total			**5964**	**57065.2**

NASA

NAS NASA Advanced Supercomputing

Space Operations and Science Grid

Collaboration between ARC CICT's Information Power Grid, MSFC Flight Projects Directorate, and JSC Mission Operations Directorate

Objectives

- Provide a fully functional Grid-enabled prototype to support distributed and secure access to Payload Operations Integration Center (POIC) ground services
- Analyze and determine feasibility of services in a Grid environment for management of ISS on-board scientific payloads

Points of Contact

Ames Research Center: Piyush Mehrotra
Marshall Space Flight Center: Robert Bradford
Johnson Space Center: Steven Gonzalez

Approach

- Create a Grid-based environment for Space Flight Operations utilizing replicated POIC ground services in a non-operational environment.
- Develop virtual organizations (VOs) along with VO-specific portals for four end-user communities
  - Payload Principal Investigators
  - Payload Control Center
  - Flight Control Center
  - Educational Outreach
- Measure the effectiveness of the Grid-enabled environment through hands-on user evaluation

Benefits

- Provide a single interface for secure and distributed access to new and current POIC ground services at significantly lower costs
- Reduce development costs for future space flight system development, replacement and upgrades.
- Leverage research and development of research enterprise and international research community

Conclusions

- The grid technology exists today to begin to implement an Earth Science Grid for the early adopters
- ARC stands ready to work with the Earth Science community to make this a reality.